Published in final edited form as:
*J Biopharm Stat*. 2013 January ; 23(1): 201–212. doi:10.1080/10543406.2013.735783.

# Medical records based post-marketing safety evaluation of rare events with uncertain status

**Hongyuan Cao**[1,*], **Lisa M. LaVange**[2], **Joseph F. Heyse**[3], **T. Christopher Mast**[4], and **Michael R. Kosorok**[2]

[1]Department of Health Studies, University of Chicago, Chicago, IL, U.S.A.

[2]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, U.S.A.

[3]Department of Biostatistics and Research Decision Sciences, Merck Research Laboratories, North Wales, PA, U.S.A.

[4]Department of Epidemiology, Merck Research Laboratories, North Wales, PA, U.S.A.

## Abstract

We develop a simple statistic for comparing rates of rare adverse events between treatment groups in post-marketing safety studies where the events have uncertain status. In this setting, the statistic is asymptotically equivalent to the logrank statistic, but the limiting distribution has Poisson and binomial components instead of being Guassian. We develop two new procedures for computing critical values, a Gaussian approximation and a parametric bootstrap. Both numerical and asymptotic properties of the procedures are studied. The test procedures are demonstrated on a post-marketing safety study of the *RotaTeq* vaccine. This vaccine was developed to reduce the incidence of severe diarrhea in infants.

### Keywords

Logrank statistic; Post-marketing safety; Poisson process; Rare events

## 1. Introduction

Rare but serious adverse events associated with vaccines (such as intussusception, for example) are difficult to detect in prelicensure studies. They require safety monitoring after introduction of the agent in large populations. Recently, the FDA mandated through its Sentinel Initiative that all approved drugs must undergo post-marketing assessment of safety. Cancer registries (see, e.g., Lucas et al. (1993)), meta-analysis of clinical trials (Berlin & Colditz (1999); Temple (1999)) and the FDA's active surveillance system MedWatch (Kessler et al. (1993)) are important and key sources of data for monitoring for rare adverse events, and there are special added levels of oversight for pediatric drug surveillance (Smith et al. (2008)).

The events reported in post-marketing safety studies based on claims data or medical records often have uncertain status, meaning that it is unclear whether a potential event is a true outcome or a reporting or classification error. To address this issue, adverse events reported as claims or medical records are frequently required to undergo adjudication by a committee to ensure that the potential events are classified correctly, with minimal bias, and

[*]hycao@uchicago.edu.

consistently across the entire study period. Adjudication committees consist of medical experts in the field of study, and the committee members are blinded to treatment assignments during the review process. The review requires more detailed information surrounding the occurrence of potential events than is typically available from claims reports or medical records. Such auxiliary data may include death certificates, hospital discharge records, outpatient medical records, and prescription claims data.

Gathering the required auxiliary data can be time-consuming and costly. Further, adjudication committees need time to review the auxiliary information and discuss their findings through committee meetings or teleconferences. Consequently, there is often a considerable delay between the time potential events are reported in the study to the time the results of the adjudication process are available for analysis. At the time of interim analysis, some potential events have been adjudicated and some have not, and thus the factuality of some of the potential endpoints is uncertain. In order to evaluate safety and report public health hazards, if any, in a timely manner, we need to be able to appropriately analyze rare events data with incomplete adjudication.

An important such example, which is in fact the main motivation for the present research, is the post-marketing safety surveillance for *RotaTeq*, a live, oral pentavalent vaccine for rotavirus sponsered by Merck Research Laboratories (Vesikari et al. (2006)). The surveillance study was a prospective observational study based on medical claims data from a large privately insured population in the United States. The study population corresponds to all infants who were enrolled in the health insurance plan within one week of birth and who were vaccinated during the course of regular clinical practice with either *RotaTeq* or diphtheriatetanus-acellular pertussis (DTaP), but not both, during the study period. Claims data were collected for the 60-day window following each dose of *RotaTeq* to determine adverse events, including intussusception. Enrollment was planned from February 2006 through December 2008. Follow-up concluded in March 2009.

The primary safety analysis for the study was a comparison of the incidence of intussusception resulting in hospitalization or emergency department visits in the 0–30 day window following vaccination (any dose) between infants receiving *RotaTeq* and infants receiving DTaP. The Safety Monitoring Committee (SMC) reviewed data quarterly to determine if the relative risk exceeded a pre-specified monitoring boundary or if other safety issues were observed. Claims-based intussusception cases were reviewed by an independent, blinded Adjudication Committee on an on-going basis and confirmed as either cases or non-cases of intussusception. During the course of the study, a portion of the medical charts were temporarily unavailable for new case review for a period of several months due to administrative reasons. This causes delays in the adjudication process and in the ability of the SMC to fully assess safety at the planned interim analysis review times, making incomplete adjudication a particular problem. Final results for the study are based on 85, 150 infants receiving at least one dose of *RotaTeq* and 62, 617 infants in the concurrent cohort receiving at least one dose of DTaP during the study period.

Although some work has been done for rare event analysis (Mast et al. (2009)), there are many unresolved issues, especially the presence of uncertain status. Important large sample approximation work for rare events include the logistic regression methodology of King & Zeng (2000) which has been implemented in commercially available software. Methodology and theory for the semi-parametric logrank test for rare survival events was developed by Buyske et al. (2000), although their results cannot deal with uncertain status of endpoints. Cook (2000) proposed statistical methods for conducting interim analyses of time-to-event data when some of the potential endpoints have not been through the adjudication process. Cook & Kosorok (2004) extended these methods, providing rigorous semi-parametric

survival analysis methodology for adjudicated endpoints, but their approach has not yet been extended to the rare event setting. Thus, new methods are crucial for monitoring studies in real time where delays in the adjudication process can occur and the event rate is rare.

In this paper, we develop and evaluate a simple test statistic that compares the event rate difference when the events have uncertain status and are partially adjudicated. We do not require the adjudication rate to be the same for the two groups. We find that, asymptotically, the intensity difference between two populations (*RotaTeq* vs DTaP) is equivalent to the logrank statistic in certain settings. We derive asymptotic distribution theory for the test statistic and provide simulations that verify the large sample properties in moderate to small samples. We then apply the new methodology to the *RotaTeq* intussusception study.

This paper is organized as follows. Section 2 describes our methods, including the main theoretical properties of the asymptotic distribution of the test statistic for rare events with incomplete adjudication and two inference procedures based on Gaussian approximation and a parametric bootstrap. Section 3 summarizes simulation results comparing different proportions of adjudication effects, type I error control and power, with respect to different numbers of true events. Section 4 applies the proposed test to data from the *RotaTeq* post-marketing safety study. Section 5 contains concluding remarks. The Appendix gives technical derivations and proofs.

## 2. Methods and test procedures

We first introduce some notation and then describe the hypothesis and test statistic that we will use. Patients are accrued randomly into two comparison arms over a study window in calendar time and followed for a claims event for time $\tau$, or the time between accrual and end of study window, which ever occurs first. Thus the relationship between accrual time and study window induces independent censoring in this setting. We assume that the length of $\tau$ is short relative to the width of the study window, and that additional independent censoring is rare, so that the probability of patients being censored before time $\tau$ is approximately zero. Thus the number of claims events, both true and false, are approximately Poisson distributed. We let $\lambda_1$ and $\lambda_2$ denote the Poisson rate for true claims cases in the two groups, $\mu_1$ and $\mu_2$ denote the two Poisson rates for false claims, and $B_1$ and $B_2$ represent the two population sizes measured in person years at risk. Specifically, the numbers of claims cases for the two groups can be modeled as $N_1 \sim \text{Poisson}((\lambda_1 + \mu_1)B_1)$ and $N_2 \sim \text{Poisson}((\lambda_2 + \mu_2)B_2)$.

Let $r_1$ and $r_2$ denote the adjudication rates for the two groups, which are assumed to be fixed and known since this is an administrative process. If we let $m_1$ and $m_2$ represent the corresponding numbers of adjudicated cases, then we can reasonably assume that $m_1 = [r_1 N_1]$ and $m_2 = [r_2 N_2]$, where $[x]$ is the smallest integer $\geq x$. Now let $A_1$ and $A_2$ represent the number of chart confirmed cases in the two groups defined as the number of events for which adjudication confirmed the claims as cases. Thus, $A_1 \sim \text{Binomial}(m_1, p_1)$ and $A_2 \sim \text{Binomial}(m_2, p_2)$, where $p_1 = \lambda_1/(\lambda_1 + \mu_1)$ and $p_2 = \lambda_2/(\lambda_2 + \mu_2)$. Our goal is to compare the intensity rates $\lambda_1$ and $\lambda_2$ for the true cases. Thus the null hypothesis is

$$H_0: \lambda_1 = \lambda_2.$$

The test statistic we propose is

$$T = \frac{N_1 A_1}{B_1 m_1} - \frac{N_2 A_2}{B_2 m_2}.$$

This can be calculated based on available data. The following theorem provides the asymptotic distribution of this test statistic:

Theorem 1: *Under* $H_0$, $\sqrt{B_1}\frac{T}{\sqrt{V(T)}} \to N(0,1)$, *as* $B_1 \to \infty$, *provided* $0 < c_1 \le \lim_{B_1\to\infty} B_2/B_1 = c \le c_2 < \infty$, *where* $V(T)=\frac{\lambda_1}{\lambda_1+\mu_1}(\lambda_1+\mu_1/r_1)+\frac{\lambda_2}{c(\lambda_2+\mu_2)}(\lambda_2+\mu_2/r_2)$.

Remark 1: Recall the logrank statistic $W=\sqrt{\frac{B_1+B_2}{B_1B_2}}\int_0^\tau \frac{Y_1(s)Y_2(s)}{Y_1(s)+Y_2(s)}[\frac{dN_1(s)}{Y_1(s)}-\frac{dN_2(s)}{Y_2(s)}]$, where $B_1$, $B_2$ are population sizes, $Y_1(s)$, $Y_2(s)$ are at risk processes and $N_1(s)$, $N_2(s)$ are counting processes for events corresponding to two groups, where the time scale is from enrollment until the earliest of $\tau$ or censoring. For extremely rare events with an approximately zero probability of being censored before $\tau$, as in our setting, $Y_i(s)/B_i \sim 1$ for all $s$; and the logrank statistic reduces to the proposed test statistic asymptotically.

Theorem 1 is a benchmark for our inference procedure. In fact, after we plug in consistent estimates of $\lambda_1$, $\mu_1$, $\lambda_2$, $\mu_2$ into $V(T)$, we can obtain a consistent variance estimate $\hat{V}(T)$ as described below, thereby enabling inference based on the proposed test statistic $T$ to be made.

Under the null hypothesis, the intensity rate is the same for the two groups, which we denote as $\lambda_0 := \lambda_1 = \lambda_2$. Below, we describe two inference procedures based on Gaussian approximation and the parametric bootstrap method:

Algorithm 1: Gaussian approximation.

Step 1. Obtain the administrative estimates of $r_1$ and $r_2$, calculate the test statistic $T$, and compute $c = B_2/B_1$ from the dataset.

Step 2. Compute parameter estimates $\widehat{\lambda_0}=\frac{N_1A_1/m_1+N_2A_2/m_2}{B_1+B_2}$, $\widehat{p_1}=\frac{A_1}{m_1}$, and $\widehat{p_2}=\frac{A_2}{m_2}$. Based on the relationship $p_i=\frac{\lambda_0}{\lambda_0+\mu_i}$, we also estimate $\widehat{\mu_i}=\frac{1-\widehat{p_i}}{\widehat{p_i}}\widehat{\lambda_0}$. $\hat{V}(T)$ is then computed by plugging in these estimates into the expression $V(T)$ in Theorem 1.

Step 3. Reject the null hypothesis if the standardized test statistic $\sqrt{B_1}\frac{|T|}{\widehat{V}(t)}$ is larger than the chosen $\alpha$-level critical value for a standard Gaussian random variable (e.g., 1.96 for $\alpha$ = 0.05).

The advantage of Gaussian approximation is that it's relatively easy to implement. But it has a potentially larger approximation error for extremely rare events, and the following parametric bootstrap method for inference can have better performance:

Algorithm 2: Parametric bootstrap method.

Step 1 is the same as in Gaussian approximation.

Step 2. Obtain parameter estimates $\hat{\lambda}_0$, $\hat{p}_1$, $\hat{p}_2$, $\hat{\mu}_1$, and $\hat{\mu}_2$, and draw random deviates $\tilde{N}_i \sim$ Pois($B_i(\hat{\lambda}_0 + \hat{\mu}_i)$), $\tilde{m}_i = r_i\tilde{N}_i$, and $\tilde{A}_i \sim$ Bin($\tilde{m}_i$, $\hat{p}_i$), $i$ = 1, 2. If any of the $\tilde{m}_i$ = 0, throw out

that samples and redraw until the number of non-rejected bootstrap samples reaches a target number (e.g., 1000). Calculate $\tilde{T}=\frac{\tilde{N}_1\tilde{A}_1}{B_1\tilde{m}_1}-\frac{\tilde{N}_2\tilde{A}_2}{B_2\tilde{m}_2}$.

Step 3. Reject the null hypothesis if the test statistic $T$ falls outside the confidence interval constructed by the $\alpha/2$ and $1-\alpha/2$ quantiles of $\tilde{T}$. For extremely rare events, we throw out samples that have value 0 for either $m_1$ or $m_2$ with an ination on the type I error control: $\tilde{\alpha}=\frac{\alpha}{P(m_1>0,m_2>0)}$. This adjustment reflects the fact that, in practice, we would only make inference based on $T$ when at least some of the data were adjudicated in both groups.

Our new procedures based on Gaussian approximation and parametric bootstrap in the presence of incomplete adjudication are in contrast with two possible procedures that did not take incomplete adjudication into account. The first possible analysis is to count only adjudicated and confirmed events and ignore the possible true events in the unadjudicated group. This "confirmed events only" (ceo) underestimates the true events rate. The second possible analysis is to count all unadjudicated events as true events in addition to the adjudicated and confirmed events. This "unrefuted events" (ure) overestimates the events rate. In fact, if we adopt approach ceo, the test statistic would be $T_{\text{ceo}}=\frac{A_1}{B_1}-\frac{A_2}{B_2}$ with null hypothesis $H_0: r_1\lambda_1 = r_2\lambda_2$. If we adopt approach ure, the test statistic would be $T_{\text{ure}}=\frac{A_1+N_1-m_1}{B_1}-\frac{A_2+N_2-m_2}{B_2}$ with null hypothesis $H_0: \lambda_1 + (1-r_1)\mu_1 = \lambda_2 + (1-r_2)\mu_2$. Neither of these alternatives can serve for our purpose.

## 3. Numerical Studies

### 3.1 Type I error control

In this section, we implement the two proposed approaches to inference with respect to different numbers of true events to see whether the type I error is preserved. The adjudication rates are set at $r_1 = 0.6$ and $r_2 = 0.5$. Population sizes are $B_1 = B_2 = 60000$. We change the true event numbers from extremely small to modest and large under the null hypothesis that the two groups have the same rate. As seen in Figure 1, when number of true events is extremely small, e.g., 6 in both groups with true event rate 0.0001, both methods are conservative with the bootstrap method being less conservative and closer to the nominal size of 0.05. When the number of true events is modest or large, the type I error control for both methods is around the nominal size 0.05. The red line denotes the nominal size of 0.05. This suggests that in practice, when we have more than 12 true events, the Gaussian approximation can provide quick and valid inference; when the number of true events is less than 12, we should consider the parametric bootstrap for more accurate type I error control. In practice, the number of true events is not available. But nevertheless, investigators usually have some rough idea based on claims data or medical records.

### 3.2 Power comparison

We next investigate the power comparing our method with the ceo and ure methods under the alternative hypothesis for different magnitudes of departure from the null hypothesis. The setup is generally the same as in the last section, except that the we change the relative difference between the two group rates from modest (50%) to large (100%). As we can see in Figure 2, our method gets better power than both ceo and ure methods. This is not surprising since these two methods ignore important features in the data and are biased. The improvement of power is especially large for modest (50%) rate difference when the true

events number is between 9 and 15. This is the right range for adverse events in post-marketing safety studies.

### 3.3 Extremely rare events

In practice, the event rates are usually extremely small. Only a couple of events occur after a few years followup based on the population size (person years at risk). In this subsection, we focus on extremely small true event numbers, from 3 to 12, and compare the performance of the two inference procedures with respect to type I error control and power. From Figure 3, again, both methods are conservative for type I error control when the number of true events is very very small and become less conservative when the number of true events increases. The bootstrap does a slightly better job when the difference in rates is 60%, but it is almost identical to the Gaussian approach when the departure from the null is large enough.

### 3.4 Adjudication effects

The different adjudication rates inevitably inuence the performance of inference procedures. We would like to see under what adjudication proportion we'll be able to get reasonable results using the method we propose. In this section, we study the effects of adjudication on type I error and power using both the Gaussian and bootstrap inference methods. Figure 4 shows that the performance under 90% adjudication is pretty close to the performance under 50% adjudication. So in practice, as long as roughly half of the outcome data are adjudicated, we would be very confident with our approach. When the adjudication rate is low, for example, only 20%, the type I error is controlled conservatively by using the Gaussian approach and asymptotically close to the nominal rate (0.05 for example) by using the bootstrap approach. When we have a relatively large number of true events, for example, 30, the power based on even 20% adjudication is pretty decent, around 70%. When true events number and the adjudication rate are both small, we almost have no detection ability —the power when number of true events is 6 and adjudication rate is 20% is less than 10%. However, if we have a fair amount of adjudication, we'll get some detection power—the power when number of true events is 6 and the adjudication rate is 50% is more than 20%. For such extremely rare events, it's comforting to see that our procedures still have some detection ability.

## 4. Data analysis

In this section, we apply both the newly developed Gaussian and parametric bootstrap inference procedures along with the ceo and ure methods to a post-marketing study of *RotaTeq* sponsored by Merck research laboratories. This vaccine is primarily developed to reduce diarrhea events due to rotavirus in infants and young children. Clinical trials have proven the effectiveness of this vaccine but its safety is yet to be evaluated after its implementation in a large population. Final results for the study are based on 85, 150 infants receiving at least one dose of *RotaTeq* and 62, 617 infants receiving at least one dose of DTaP during the study period. Infants receiving at least one dose of *RotaTeq* contributed $B_1$ = 17, 433 person-years of follow-up for adverse events occurring in the 0–30 day window following any dose. The DTaP comparison group contributed $B_2$ = 12, 339 person-year of follow-up for the same window. Claims-based intussusception cases were reviewed by an independent, blinded Adjudication Committee on an on-going basis and confirmed as either cases or non-cases. Table 1 summarizes the dataset. For the 0 – 30 window, one chart was not available for a *RotaTeq* case and one chart was not available for the DTaP suspected case in the 0 – 60 window (outside 0 – 30). Adjudication numbers did not change between interim 1 and interim 2. We calculate p-values based on the ceo method, the ure method and the two new inference procedures we proposed. The final results are very similar with each other since at most one case was not adjudicated by the final stage. We did not see a

significantly greater number of intussusception cases in infants taking *RotaTeq* compared with infants taking DTaP.

## 5. Summary and Discussion

In this paper, we present inference procedures for analyzing rare events with uncertain status using both a Gaussian approximation and a parametric bootstrap. These methods can potentially be extended to the multi-category primary event situation. We provide unbiased estimates of the true event rates and perform formal hypothesis testing with incomplete adjudication data. Our method works best when at least half of the cases are adjudicated (Figure 4). It is based on the assumption that the probabilities of true events rates are the same for the un-adjudicated and adjudicated cases. The importance of the new methods lie in the fact that we can identify signals early enough with confidence. This can potentially save lives if we see any significant difference in a timely manner. If we wait until all the data are adjudicated, this could slow down the surveillance process and possibly adversely affect drug safety. We observed that the two approaches perform comparably when the number of true events is greater than 12 with sample size 60000 person years in both groups, while the parametric bootstrap provides a less conservative type I error and is therefore preferred for extremely small numbers of true events based on our simulation study. When the relative difference between the group rates is larger than 50%, we obtain slightly better power when using the bootstrap method. We applied both methods to real data from two of the planned interim analyses and the final analysis of a post-marketing surveillance study. The results for the two inferential approaches were similar. In order to use these inferential procedures, we must have a positive number of adjudicated events. We did not take into account reporting bias in our analysis. This could potentially impact on the analysis results and be pursued as a future research topic.

## Acknowledgments

The authors are grateful to the funding provided by Merck Research Laboratories. The last author was also supported in part by NCI grant CA142538.

## Appendix

In this section, we prove Theorem 1 by deriving the asymptotic distribution of $\frac{N_1A_1}{B_1m_1}$. The basic structure we have is that $N_1 \sim \text{Pois}(B_1(\lambda_1 + \mu_1))$, $m_1|N_1 = r_1N_1$, $A_1|m_1$, and $N_1 \sim Bin(m_1, \frac{\lambda_1}{\lambda_1+\mu_1})$. Also,

$$E\left(\frac{N_1A_1}{B_1m_1}\right) = E\left[E\left(\frac{N_1A_1}{B_1m_1}|N_1, m_1\right)\right] = E\left[\frac{N_1\lambda_1/(\lambda_1+\mu_1)}{B_1}\right] = \lambda_1,$$

$$Var\left(\frac{N_1A_1}{B_1m_1}\right) = Var\left[E\left(\frac{N_1A_1}{B_1m_1}|N_1, m_1\right)\right] + = E\left[Var\left(\frac{N_1A_1}{B_1m_1}|N_1, m_1\right)\right] = Var\left[\frac{N_1\lambda_1/(\lambda_1+\mu_1)}{B_1}\right] + E\left[\frac{N_1^2\lambda_1\mu_1/(\lambda_1+\mu_1)^2}{B_1^2m_1}\right] = \frac{\lambda_1^2}{B_1(\lambda_1+\mu_1)} + \frac{\lambda_1\mu_1}{B_1r_1(\lambda_1+\mu_1)}.$$

In fact, given the assumption that $m_1 = r_1N_1$, the test statistic $T = \frac{A_1}{B_1r_1}$, where $B_1$ and $r_1$ are fixed and known. The only random variable involved is $A_1 = \sum_{i=1}^{m_1} X_i$, where $X_i$ are i.i.d. binomial random variables with success probability $p_1 = \frac{\lambda_1}{\lambda_1+\mu_1}$, and $m_1$ is a random variable as well. Through the characteristic function method, we can show that

$$\sqrt{B_1}\left(\frac{N_1A_1}{B_1m_1} - \lambda_1\right) \to N\left(0, \frac{\lambda_1}{\lambda_1+\mu_1}(\mu_1/r_1+\lambda_1)\right).$$

The proof is somewhat involved and we refer to Robbins (1948) for the details. Note that under the null hypothesis $\lambda_1 = \lambda_2 = \lambda_0$ and the two groups are independent of each other. This completes the proof of Theorem 1.

## References


Berlin J, Colditz G. The role of meta-analysis in the regulatory process for food, drugs and devices. Journal of the American Medical Associaiton. 1999; 28:830–834.

Buyske S, Fagerstrom R, Ying Z. A class of weighted log-rank tests for survival data when the event is rare. J. Am. Statist. Assoc. 2000; 95:249–258.

Cook T. Adjusting survival analysis for the presense of non-adjudicated study events. Controlled Clinical Trials. 2000; 21:208–222. [PubMed: 10822119]

Cook T, Kosorok M. Analysis of time-to-event data with incomplete event adjudication. J. Am. Statist. Assoc. 2004; 99:1140–1152.

Kessler D, Natanblut S, Kennedy D, et al. A new approach to reporting medication and device adverse effects and product problems. Journal of the American Medical Association. 1993; 269:2765–2768. [PubMed: 8492403]

King G, Zeng L. Logistic regression in rare events data. The Global Burden of Disease 2000 in Aging Populaitons, Research Paper No.2 Harvard University. 2000

Lucas T, Beresford J, Vessey M. Delay and completeness of cancer registration in a post-marketing surveillance study of cimetidine. Pharmacoepidemiology and Drug Safety. 1993; 2:201–203.

Mast T, Loughlin J, Wang F, Doherty M, Seeger J. Us postlicensure active surveillance safety study of rotateq™, oral pentavalent rotavirus vaccine (rv5). 2009 Unpublished manuscript.

Robbins H. The asymptotic distribution of the sum of a random number of random variables. Bull. Amer. Math. Soc. 1948; 54:1151–1161.

Smith P, Benjamin D, Murphy M, Johann-Liang R, Lyasu S, Gould B, Califf RM, Li J, Rodriguez W. Safety monitoring of drugs receiving pediatric marketing exclusivity. Pediatrics. 2008; 122:e628–e633. [PubMed: 18762496]

Temple R. Meta-analysis and epidemiologic studies in drug development and postmarketing surveillance. Journal of the American Medical Association. 1999; 281:841–844. [PubMed: 10071007]

Vesikari T, Matson D, Dennehy P, et al. Safety and efficacy of a pentavalent human-bovine(wc3) reassortant rotavirus vaccine. The New England Journal of Medicine. 2006; 354:23–33. [PubMed: 16394299]

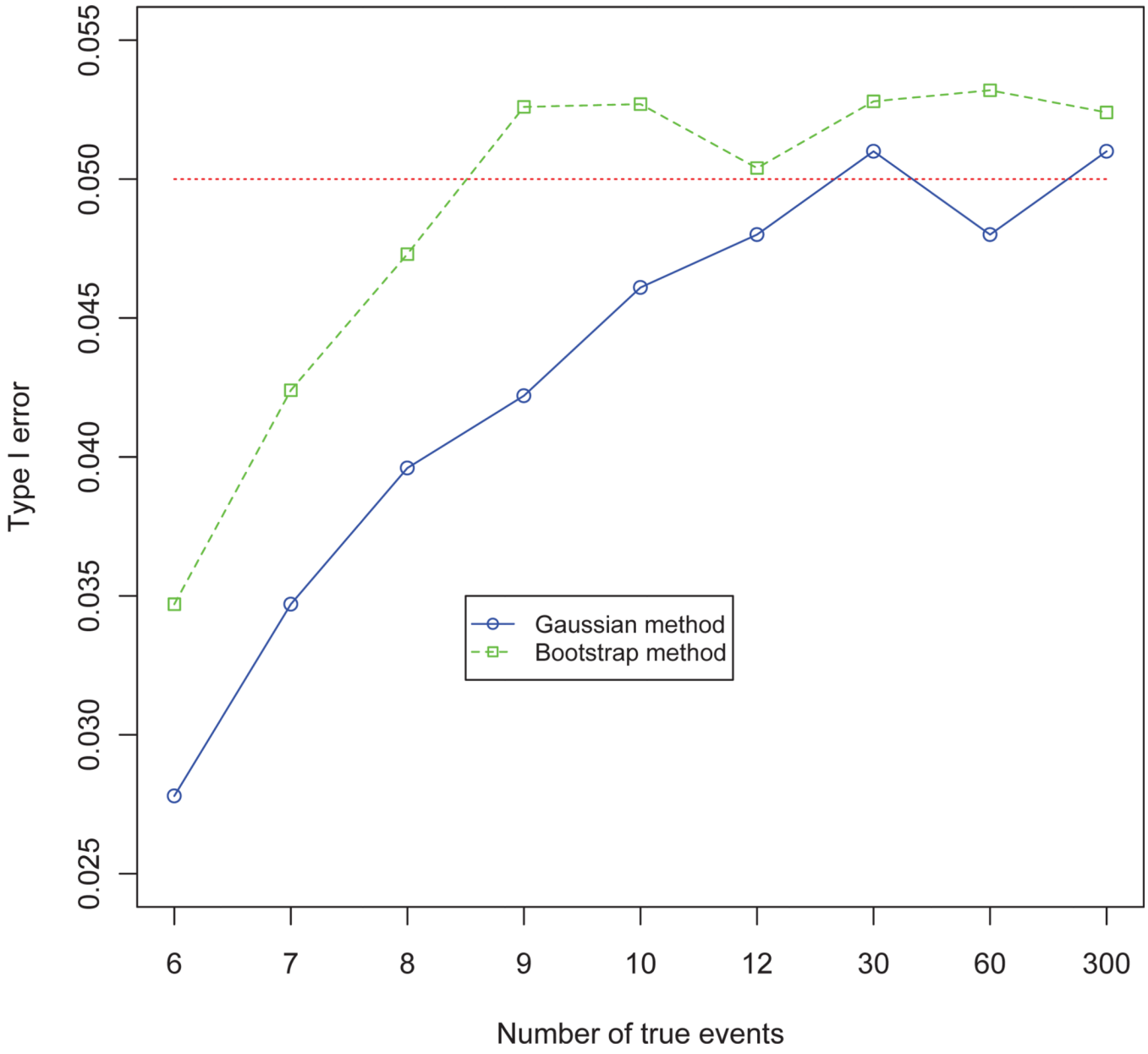


**Figure 1.**
Type I error control with different number of true events

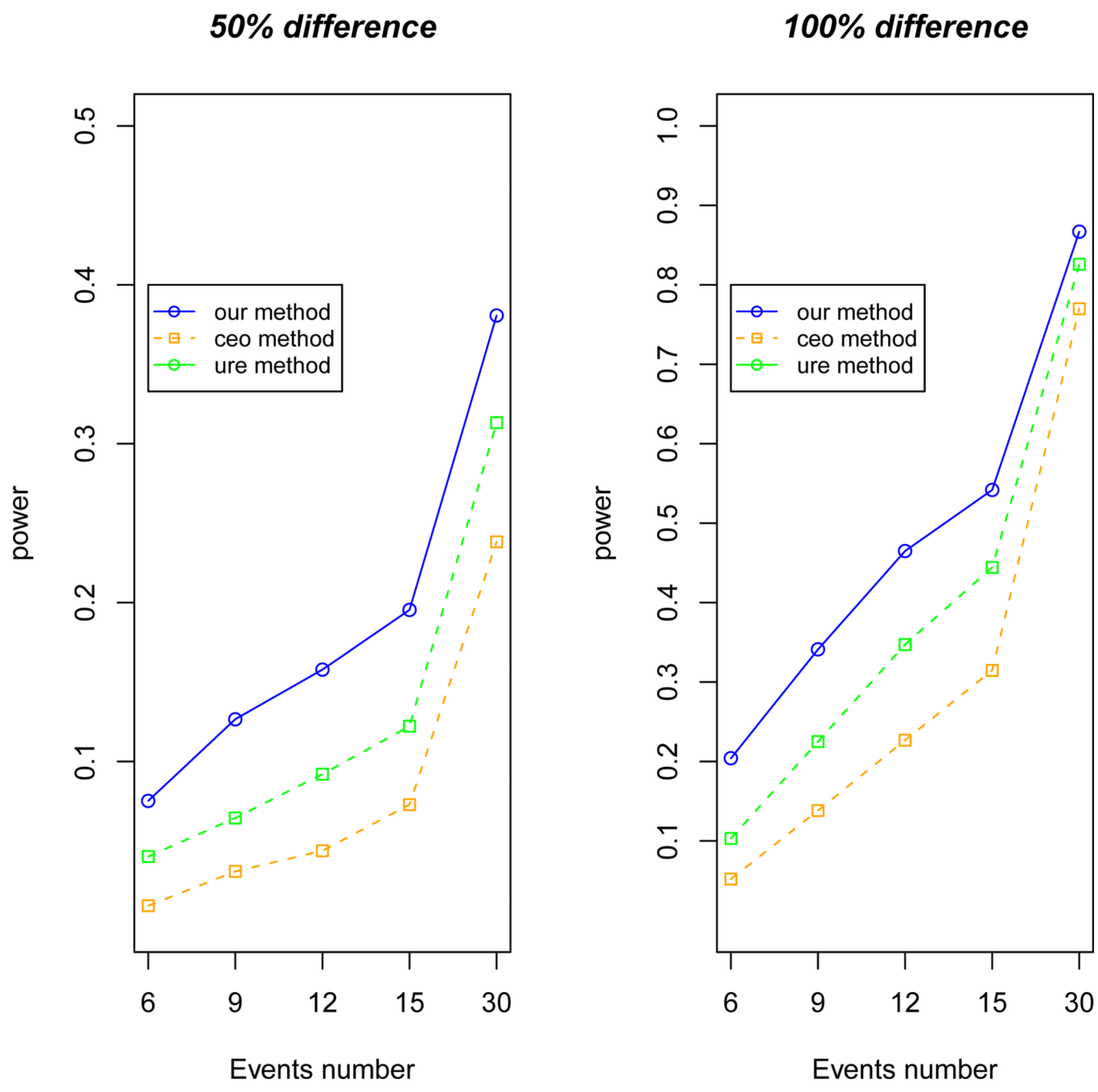


**Figure 2.**
Comparison of power among our method, ceo method and ure method

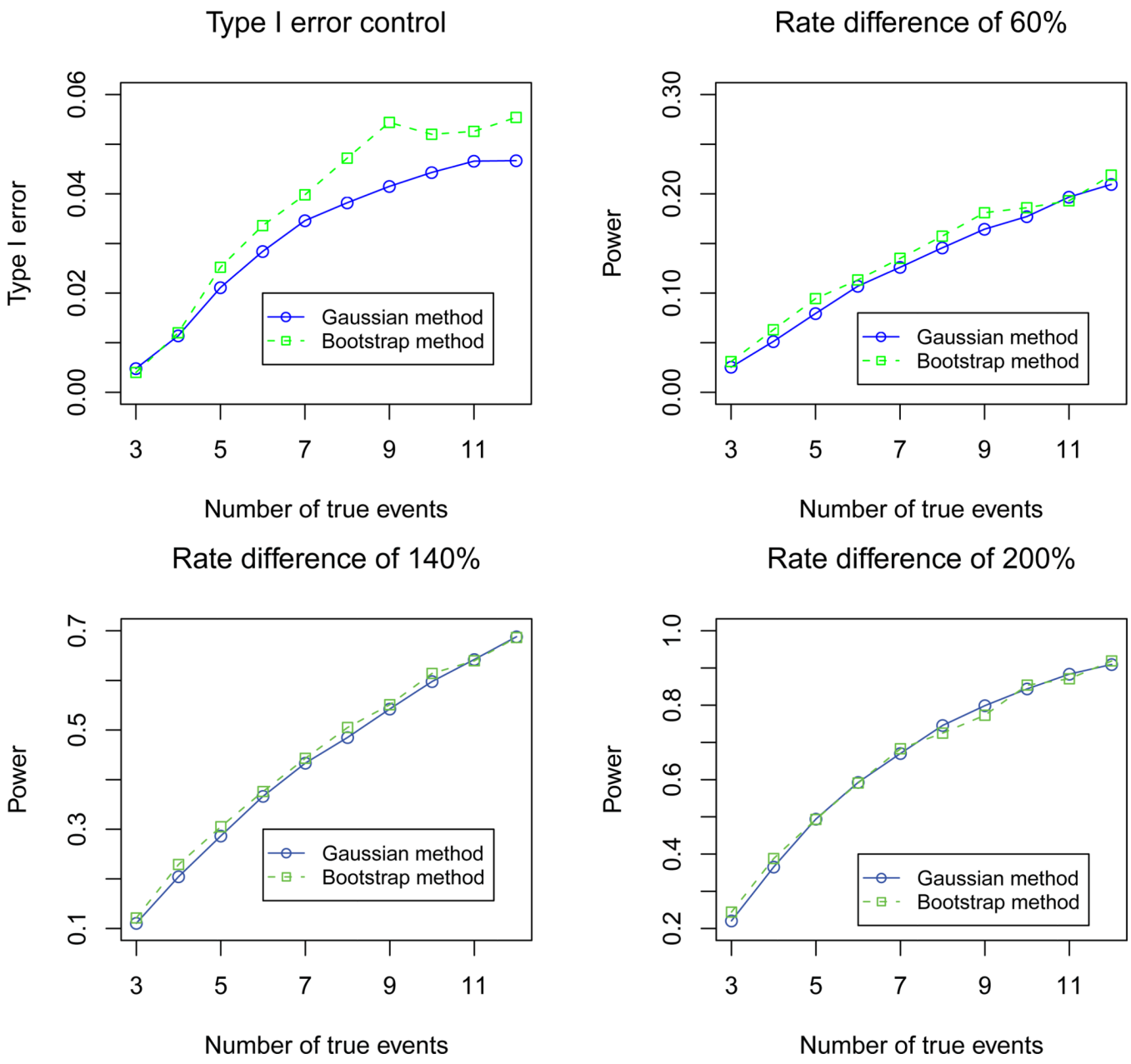


**Figure 3.**
Type I error and power for extremely rare events

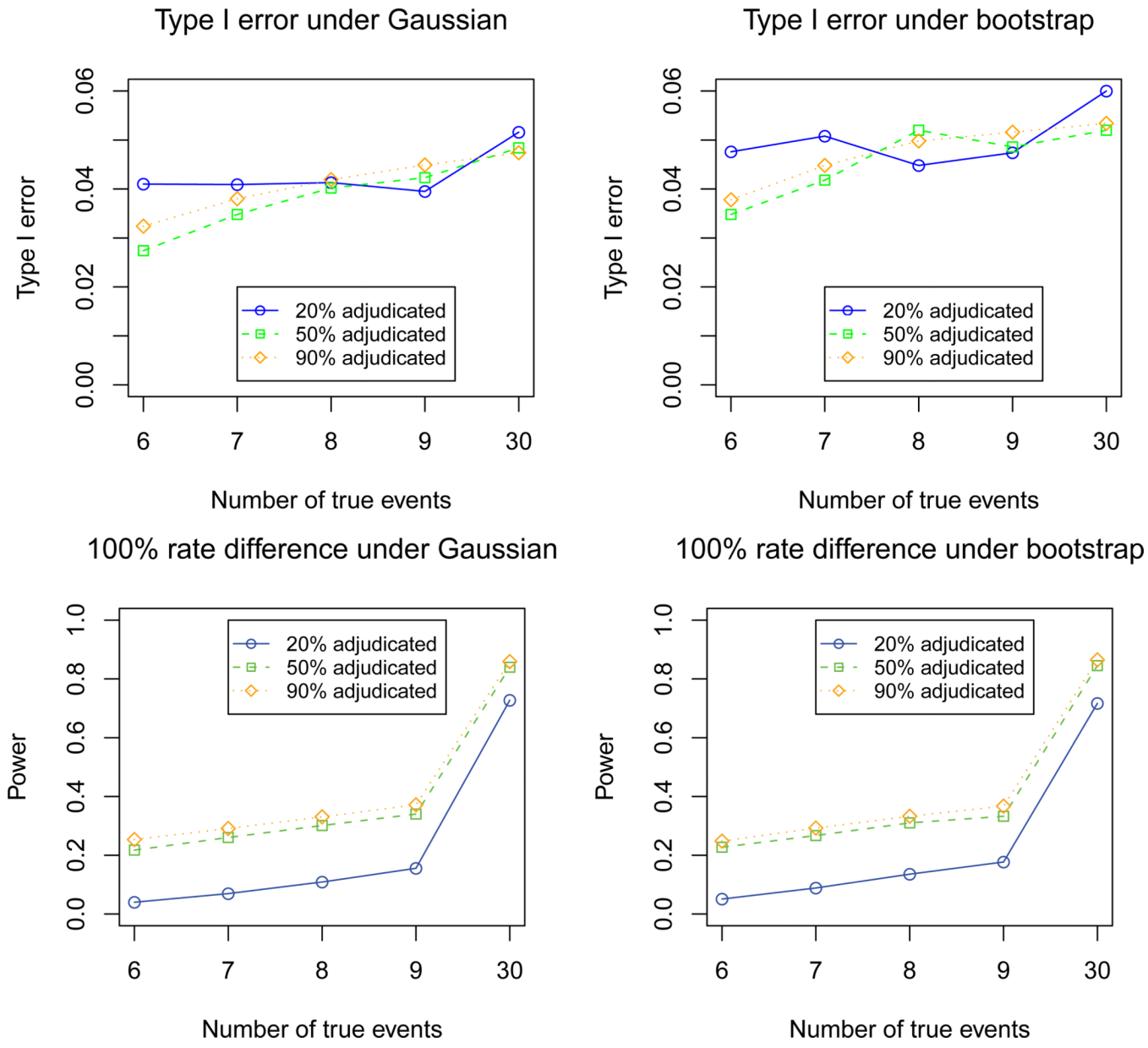


**Figure 4.**
Impacts of different adjudication rates

**Table 1**

Data Summary

| | 0 – 30 window | | | 0 – 60 window | | |
|---|---|---|---|---|---|---|
| | **interim *I*** | **interim *II*** | **final** | **interim *I*** | **interim *II*** | **final** |
| B1 | 13, 223 | 13, 641 | 17, 433 | 24, 887 | 25, 674 | 32, 799 |
| B2 | 10, 928 | 11, 098 | 12, 339 | 20, 578 | 20, 895 | 23, 211 |
| N1 | 13 | 13 | 15 | 19 | 20 | 22 |
| N2 | 5 | 6 | 9 | 12 | 13 | 17 |
| A1 | 2 | 2 | 6 | 4 | 4 | 11 |
| A2 | 4 | 4 | 5 | 5 | 5 | 6 |
| m1 | 6 | 6 | 14 | 9 | 9 | 21 |
| m2 | 5 | 5 | 9 | 8 | 8 | 16 |
| p-value(gaussian) | 0.891 | 0.690 | 0.875 | 0.903 | 0.842 | 0.623 |
| p-value(bootstrap) | 0.872 | 0.704 | 0.880 | 0.922 | 0.896 | 0.604 |
| p-value(ceo) | 0.381 | 0.383 | 0.827 | 0.635 | 0.623 | 0.682 |
| p-value(ure) | 0.257 | 0.464 | 0.987 | 0.534 | 0.607 | 0.680 |